



PATENT

MS154753.01

CERTIFICATE OF MAILING

I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Mail Stop Appeal Brief-Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Date: 7-15-04

Himanshu S. Amin
Himanshu S. Amin

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of:

Applicant(s): Eric W. Nace *et al.*

Serial No: 09/633,326

Filing Date: August 7, 2000

Examiner: Yemane M. Gerezgiher

Art Unit: 2144

Title: SYSTEM AND METHOD PROVIDING CONTINUAL RATE REQUESTS

**Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450**

APPEAL BRIEF

Dear Sir:

Applicants submit this Appeal Brief in triplicate in connection with an appeal of the above-identified patent application. Please charge \$330.00 for the fee associated with this brief to Deposit Account No. 50-1063 [MSFTP124US].

07/21/2004 RMEBRAHT 00000017 501063 09633326

01 FC:1402 330.00 DA

I. Real Party in Interest (37 C.F.R. §1.192(c)(1))

The real party in interest in the present appeal is Microsoft Corporation, the assignee of the present application.

II. Related Appeals and Interferences (37 C.F.R. §1.192(c)(2))

Appellants, appellants' legal representatives, and/or the assignee of the present application are not aware of any appeals or interferences which will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. Status of Claims (37 C.F.R. §1.192(c)(3))

Claims 1-21 are currently pending in the subject application and are presently under consideration. The rejection of claims 1-21 is appealed.

IV. Status of Amendments (37 C.F.R. §1.192(c)(4))

No claim amendments have been entered subsequent the Final Office Action.

V. Summary of Invention (37 C.F.R. §1.192(c)(5))

The subject invention relates to systems and methods that provide a consistent server load, which can be adjusted and controlled to determine server capacity. (*See* pg. 1, ln. 8-10). A user can specify a rate of a continual stream of requests, such as HTTP requests, sent from a client to a server. (*See* pg. 2, ln. 23-25). The subject invention can provide the continual stream of requests by utilizing a closed loop system wherein feedback is employed within the framework of a queuing architecture to control a desired network load. (*See* pg. 4, ln. 19-22). For example, the queuing architecture can comprise a queue for processing the requests, a component for sending/receiving the requests, a feedback loop for controlling the desired rate of the requests, and a scheduler for determining a continual rate of requests for an upcoming period. (*See* pg. 3, ln. 4-7). The feedback loop can be provided with a target rate per second and an actual rate per second, and can determine an error signal corresponding to the difference between the actual and target rates. (*See* pg. 7, ln. 23-27). The user specified consistent rate of requests can be utilized, for example, to determine whether particular code and/or applications operate under load, whether any negative results are produced from the load, whether breaking points are

produced from excessive load, and/or whether new applications and/or hardware have the capacity to serve the desired load. (See pg. 2, ln. 27 – pg. 3, ln. 2).

VI. Statement of the Issues (37 C.F.R. §1.192(c)(6))

Whether claims 1, 4-12, 17, and 18 are unpatentable under 35 U.S.C. §102(a) as being anticipated by *Banga et al.* (“Measuring the Capacity of a Web Server”).

Whether claims 13-16 are unpatentable under 35 U.S.C. §103(a) over *Banga et al.* (“Measuring the Capacity of a Web Server”) in view of *Yu* (U.S. 6,078,943).

Whether claims 2 and 3 are unpatentable under 35 U.S.C. §103(a) over *Banga et al.* (“Measuring the Capacity of a Web Server”) in view of *Dantressangle* (U.S. 6,446,120).

Whether claims 19-21 are unpatentable under 35 U.S.C. §103(a) over *Banga et al.* (“Measuring the Capacity of a Web Server”) in view of what would have been obvious to one having ordinary skill in the art at the time the invention was made.

VII. Grouping of Claims (37 C.F.R. §1.192(c)(7))

For the purposes of this appeal only, the claims are grouped as follows:

Claims 1-21 stand or fall together.

VIII. Argument (37 C.F.R. §1.192(c)(8))

A. Rejection of Claims 1, 4-12, 17, and 18 Under 35 U.S.C. §102(a)

Claims 1, 4-12, 17, and 18 stand rejected under 35 U.S.C. §102(a) as being anticipated by *Banga et al.* (“Measuring the Capacity of a Web Server”). It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. *Banga et al.* does not teach or suggest *each and every limitation* of the claimed invention.

i. *Applicable Law*

A single prior art reference anticipates a patent claim only if it expressly or inherently describes *each and every limitation* set forth in the patent claim. *Trintec Industries, Inc., v. Top-U.S.A. Corp.*, 295 F.3d 1292, 63 U.S.P.Q.2D 1597 (Fed. Cir. 2002). “A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described in a single prior art reference.” *Verdegaal Bros. v. Union Oil Co. of*

California, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). "The identical invention must be shown in as complete detail as is contained in the ... claim." *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989) (emphasis added).

ii. ***Banga et al. does not teach or suggest each and every limitation set forth in claims 1, 4-12, 17, and 18. Therefore, Banga et al. does not anticipate claims 1, 4-12, 17, and 18.***

The present invention relates to utilizing a consistent server load, which can be adjusted and controlled, to determine server capacity. (See pg. 1, ln. 8-10). A user can specify a rate of a continual stream of requests sent from a client to a server. (See pg. 2, ln. 23-25). In particular, as recited in independent claim 1 (and similarly in independent claims 17 and 18), the *client provid[es] a desired rate of requests by calculating an actual rate of requests being generated and adjusting the actual rate to within a predetermined range of the desired rate such that a continual rate of requests are provided to the server in order to facilitate determining server capacity*. *Banga et al.* fails to teach or suggest such claimed aspects of the subject invention. Instead, *Banga et al.* discloses a method for Web traffic generation that can generate bursty traffic, with peak loads that exceed the capacity of the server. (See Abstract).

More particularly, one novel aspect of applicants' claimed invention is that a consistent and predictable load for a server can be established to determine the capacity of a given server. (See pg. 2, ln. 25-27). By way of example, utilization of a predetermined load at a consistent rate for the server facilitates determination of whether particular code and/or applications operate under load, whether any negative results are produced from the load, whether breaking points are produced from excessive load, and/or whether new applications and/or hardware have the capacity to serve the desired load. (See pg. 2, ln. 27 – pg. 3, ln. 2).

Banga et al. does not teach or suggest that *a continual rate of requests are provided to the server* as recited in independent claim 1 (and similarly in independent claims 17 and 18). *Banga et al.* relates to bursty requests rates, "including peak loads that exceed capacity of the server." (See pg. 2, first partial paragraph). The Advisory Action dated June 4, 2004 notes "generating HTTP requests at a certain rate and with a certain request distribution." (See Advisory Action, pg. 2). Applicants' representative submits that this statement is taken out of

context. More particularly, Banga *et al.* discloses that an S-Client consists of a pair of processes, one of which is “the connection establishment process [that] is responsible for generating HTTP requests at a certain rate and with a certain request distribution.” (See pg. 5, full paragraph 5). Additionally, a number of S-Clients are employed. Thus, each S-Client is generating HTTP requests at a certain rate and with a certain distribution. This is similar to the concurrent connection model, which does not provide a controllable amount of stress for a server. (See pg. 1, ln. 23-29). Therefore, Banga *et al.* fails to teach or suggest that ***a continual rate of requests are provided to the server*** as claimed.

Moreover, Banga *et al.* fails to teach or suggest ***the client providing a desired rate of requests*** as recited in independent claim 1 (and similarly in independent claims 17 and 18). Banga *et al.* instead discloses utilization of a set of client machines, where each client runs a number of S-Client processes. (See pg. 5, full paragraph 4). Each S-Client comprises a connection establishment process, which generates HTTP requests at a certain rate and with a certain request distribution. (See pg. 5, full paragraph 5). The techniques utilized in Banga *et al.* generate request rates beyond the capacity of the server. (See pg. 6, full paragraph 3). Banga *et al.* is silent regarding coordination of the requests from each of the S-Clients to provide a desired rate of requests.

It is contended in the Advisory Action dated June 4, 2004 that:

[W]hen a client generated an HTTP request to the server, it is broadly interpreted to mean it's desired request where the desired request is the capacity of the server. Having said that, the scalable request generating means of Banga adjusts the request rate in accordance with the capacity of the server that is being stress tested. (See Page 5, Section 4). Once the server can't longer handle the generated requests, a capacity of a server is determined ('... 130 requests per second, which is the capacity of our server for this request size ... the request rate remains nearly constant at the capacity of the server')

(See pg. 2). Applicants respectfully disagree with such assertion. Banga *et al.* notes that in a conducted experiment, the simple scheme generated no more than about 130 requests per second, which was the capacity for the server under the employed experimental conditions. (See pg. 8, full paragraph 3). Thus, the 130 requests per second are not a ***desired rate of requests***; instead, the 130 requests per second are the maximum capacity for the server employed in the Banga *et*

al. experimentation. The maximum capacity of a server is not a rate that can be defined or specified by a user. Therefore, Banga *et al.* does not teach or suggest ***the client providing a desired rate of requests.***

Furthermore, Banga *et al.* is silent regarding providing a desired rate of requests by ***calculating an actual rate of requests being generated and adjusting the actual rate to within a predetermined range of the desired rate such that a continual rate of requests are provided to the server*** as recited in independent claim 1 (and similarly in independent claims 17 and 18). Banga *et al.* merely discloses that a constant think time can be utilized to achieve a certain constant request rate from an S-Client. Banga *et al.* does not teach or suggest ***calculating an actual rate of requests or adjusting the actual rate to within a predetermined range of the desired range.***

The Advisory Action dated June 4, 2004 contends that “the specific server capacity (130/second) described must have been (inherently disclosed) calculated based on the desired client rate of requests (scalable requests generated) and actual generated requests.” (See pg. 2). Applicants respectfully disagree.

In a patent case, under the doctrine of inherency, if an element is not expressly disclosed in a prior art reference, the reference will still be deemed to anticipate a subsequent claim if the missing element is necessarily present in the thing described in the reference, and that it would be so recognized by persons of ordinary skill. Inherent anticipation requires that the missing descriptive material is ***necessarily present, not merely probably or possibly present***, in the prior art. *Rosco, Inc. v. Mirror Lite Co.*, 304 F.3d 1373, 64 USPQ2d 1676 (Fed. Cir. 2002) (emphasis added).

Applicants’ representative respectfully submits that Banga *et al.* does not teach or suggest providing a desired rate of requests as noted *supra*. Moreover, providing a desired rate of requests by calculating an actual rate of requests being generated and adjusting the actual rate to within a predetermined range of the desired rate is not necessarily present in Banga *et al.* Thus, such an element is not inherent in Banga *et al.*

In view of at least the above, it is readily apparent that Banga *et al.* does not anticipate or suggest the subject invention as recited in independent claims 1, 17, and 18 (and claims 4-12 which respectively depend there from). This rejection should be withdrawn.

B. Rejection of Claims 13-16 Under 35 U.S.C. §103(a)

Claims 13-16 stand rejected under 35 U.S.C. §103(a) as being unpatentable over *Banga et al.* ("Measuring the Capacity of a Web Server") in view of Yu (US 6,078,943). It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. *Banga et al.* and Yu, individually or in combination, do not teach or suggest each and every element set forth in the subject claims.

Yu does not make up for the aforementioned deficiencies of *Banga et al.* with respect to independent claim 1 (which claims 13-16 directly or indirectly depend from). Yu does not teach or suggest a client ... providing a *desired rate of requests* by calculating an *actual rate* of requests being generated and *adjusting the actual rate to within a predetermined range* of the desired rate such that a continual rate of requests are provided to the server. Instead, Yu discloses utilizing an arbiter, which can assign clients to servers or assign a valid time interval to each mapping request based on network load and/or capacity parameters. (See Abstract). Therefore, the subject invention as recited in claims 13-16 is not obvious over the combination of *Banga et al.* and Yu. Accordingly, withdrawal of this rejection is respectfully requested.

C. Rejection of Claims 2 and 3 Under 35 U.S.C. §103(a)

Claims 2 and 3 stand rejected under 35 U.S.C. §103(a) as being unpatentable over *Banga et al.* ("Measuring the Capacity of a Web Server") in view of Dantressangle (US 6,446,120). It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. *Banga et al.* and Dantressangle, individually or in combination, do not teach or suggest each and every element set forth in the subject claims.

Dantressangle does not make up for the aforementioned deficiencies of *Banga et al.* with respect to independent claim 1 (which claims 2 and 3 directly or indirectly depend from). Dantressangle does not teach or suggest a client ... providing a *desired rate of requests* by calculating an *actual rate* of requests being generated and *adjusting the actual rate to within a predetermined range* of the desired rate such that a continual rate of requests are provided to the server. Instead, Dantressangle discloses creating one or more virtual browsers at a client computer for transmitting commands to the server computer. (See Abstract). Therefore, the

subject invention as recited in claims 2 and 3 is not obvious over the combination of Banga *et al.* and Dantressangle. Accordingly, withdrawal of this rejection is respectfully requested.

D. Rejection of Claims 19-21 Under 35 U.S.C. §103(a)

Claims 19-21 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Banga *et al.* ("Measuring the Capacity of a Web Server") in view of what would have been obvious to one having ordinary skill in the art at the time the invention was made. It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. Banga *et al.* does not teach or suggest each and every element of independent claim 18 (which claims 19-21 directly or indirectly depend from) as discussed *supra*. Therefore, the subject invention as recited in claims 19-21 is not obvious over Banga *et al.* Accordingly, withdrawal of this rejection is respectfully requested.

IX. Conclusion

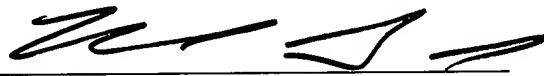
The present application is believed to be in condition for allowance, in view of the above comments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063.

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicants' undersigned representative at the telephone number listed below.

Respectfully submitted,

AMIN & TUROCY, LLP



Himanshu S. Amin

Reg. No. 40,894

AMIN & TUROCY, LLP
24TH Floor, National City Center
1900 E. 9TH Street
Cleveland, Ohio 44114
Telephone (216) 696-8730
Facsimile (216) 696-8731

X. Appendix of Claims (37 C.F.R. § 1.192(c)(9))

1. (Original): A system for determining capacity of a server, comprising:
a client for generating a plurality of requests to the server, the client providing a desired rate of requests by calculating an actual rate of requests being generated and adjusting the actual rate to within a predetermined range of the desired rate such that a continual rate of requests are provided to the server in order to facilitate determining server capacity.
2. (Original): The system of claim 1 further including a control input for adjusting the desired rate of requests.
3. (Original): The system of claim 2 wherein the control input provides the desired rate of requests.
4. (Original): The system of claim 2 wherein capacity planning is provided by monitoring performance metrics on the server while adjusting the desired rate of requests.
5. (Original): The system of claim 1 wherein capacity planning is provided automatically by monitoring performance feedback from the server.
6. (Original): The system of claim 5 wherein capacity planning is determined by automatically adjusting the desired rate of requests and comparing the performance feedback to predetermined thresholds.
7. (Original): The system of claim 1 further including a data store for holding a plurality of requests.
8. (Original): The system of claim 7 further including a queuing mechanism for retrieving and sorting requests from the data store.

-
9. (Original): The system of claim 8 further including a queue for storing sorted requests from the queuing mechanism.
 10. (Original): The system of claim 9 wherein the requests are sorted according to the criteria of time to execute.
 11. (Original): The system of claim 10 wherein the requests are HTTP requests.
 12. (Original): The system of claim 9 further including a component for retrieving requests from the queue and sending the requests to the server.
 13. (Original): The system of claim 1 further including a scheduler for determining how many requests to generate for an upcoming period.
 14. (Original): The system of claim 13 further including a feedback loop for controlling the desired rate of requests.
 15. (Original): The system of claim 14 wherein the feedback loop determines an error signal that is provided to the scheduler for controlling the desired rate of requests.
 16. (Original): The system of claim 13 wherein the scheduler is activated during a current time period to schedule requests for an upcoming time period.
 17. (Original): A system for determining capacity of a server, comprising:
means for generating a plurality of requests to the server, said means providing a desired rate of requests by calculating actual rate of requests being generated and adjusting the actual rate to within a predetermined range of the desired rate such that a continual rate of requests are provided to the server in order to facilitate determining server capacity.

18. (Original): A methodology for generating a continual stream of network requests comprising:

scheduling requests for an upcoming period of time;

sampling actual requests per second;

determining if the actual requests per second are below a target requests per second; and

increasing the actual requests per second in the upcoming period if the actual requests per second are below the target requests per second.

19. (Original): The methodology of claim 18 wherein the step of determining if actual requests per second are below the target requests per second is determined by performing a subtraction.

20. (Original): The methodology of claim 18 wherein the actual requests per second are decreased if the actual requests per second are above the target requests per second.

21. (Original): The methodology of claim 18 wherein the actual requests per second are maintained if the actual requests per second are equal to the target requests per second.